



Project StORe

- expectations, a solution
and some predicted impact from
opening up the research data portfolio

Graham Pryor



StORe guide

- Libraries and the research agenda
- The StORe project
 - Objectives
 - Results from a survey
- Support and self-determination
- The challenge for librarians?

Libraries without walls 7



Libraries without walls 7

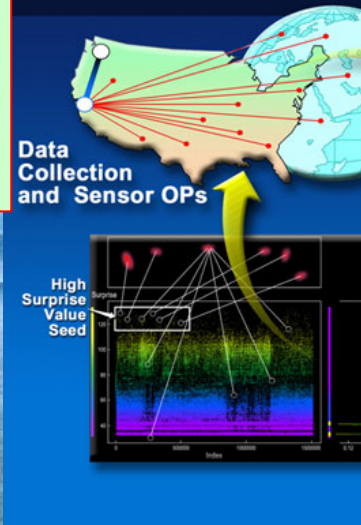


Libraries without walls 7

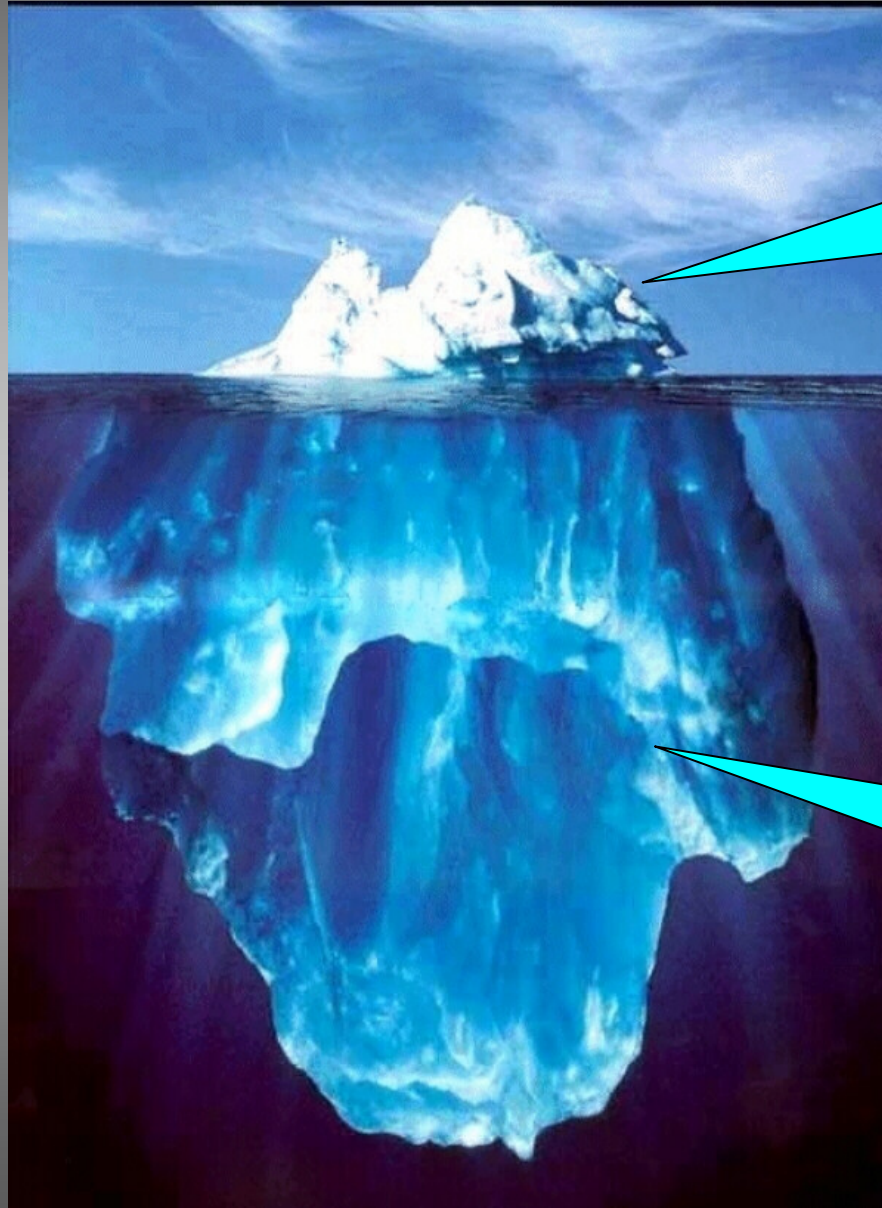


Challenges in Internet Anomaly Detection

**£3,000,000,000 current
annual investment – UK
Research Councils**



Libraries without walls 7



Scholarly publications – the visible research output

Research data – the hidden asset (just melting away?)

Libraries without walls 7



The pencil is mightier than the pen

- Robert M Pirsig, *Lila – an inquiry into morals*

- Science as provisional
- Contains a mechanism “whereby new dynamic insight can wipe out old static patterns without wiping out science itself”
- Enables continuous evolutionary growth



~~The pencil is mightier than the pen~~

The computer is mightier than the
printing press

So, the next stage in the evolutionary
growth of ~~science~~ librarians will be...?



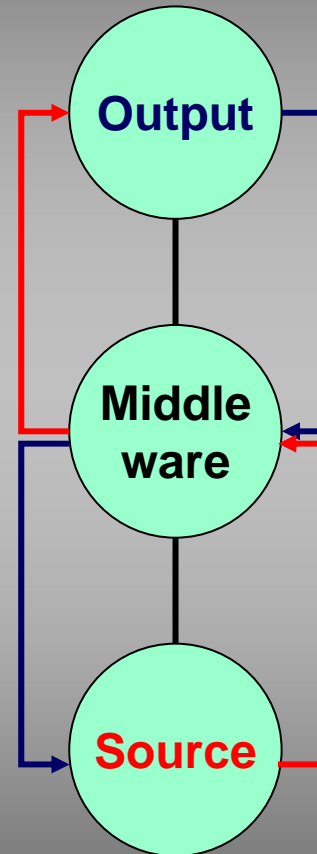
Political landscape

- Open access movement – the Berlin, Budapest and other declarations
- Wellcome Trust
 - Funding linked to commitment to deposit
- UK Research Councils – data policies
 - Publicly funded research data a public good
 - Appropriate curation throughout the lifecycle
 - Need for high quality metadata



The StORe mission

- Seamless transport from research data to research publications and vice versa
- Bi-directional links proven in social science e-research but capable of export to other disciplines



Libraries without walls 7



SuperCOSMOS Science Archive

SSA - SuperCOSMOS Science Archive

SSA Home

Data Overview

Schema browser

Data access

Radial

Menu query

Freeform SQL

CrossID

Cookbook

Personal SSA

Q&A

Release History

Downtime

Links

Credits

The SuperCOSMOS Science Archive (SSA) is a digital repository of astronomical survey plates. At around 2.5 terabytes over 1 billion multi-color wavebands (BRI), with

Access to the data has been based on the same underlying platform allows users to

A short description of the [Schema Browser](#).

Users wishing to access

Edinburgh Research Archive : Home - Microsoft Internet Explorer

Address <http://www.era.lib.ed.ac.uk/index.jsp>

Edinburgh Research Archive

Search ERA

Go

Advanced Search >>

Not logged in ([Login](#))

Edinburgh Research Archive >

Welcome to the Edinburgh Research Archive

Welcome to ERA, a digital repository of research output from The University of Edinburgh. Contained here are full-text digital Theses and Dissertations, book chapters, journal pre-prints and peer-reviewed journal reprints.

Search

Enter some text in the box below to search ERA.

Go

Or go to the [Advanced Search](#)

Browse

- [Communities & Collections](#)
- [by Title](#)
- [by Author](#)
- [by Date](#)

Sign on to:

- [Receive email updates](#)
- [My ERA](#)
authorized users
- [Edit Profile](#)

Depositing Items

Learn how to deposit your items in ERA: Read the [Deposit Guide](#)

Copyright

Not sure of the Publishers copyright policy to archiving your journal articles online? Use these sites to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.

By Journal Publisher: <http://www.sherpa.ac.uk/romeo.php>

By Journal Title: <http://romeo.eprints.org/>

Similarly, if you are submitting an electronic thesis to ERA you may be worried that you are violating publishers' publication policies for future journal articles you may wish to publish. You can check the database of links to publishers' journal policies called the Academic Journal Policy Database (AJPD): <http://www.etd.uc.edu/journal/>

Mediated Deposit

Are you or your department too busy to archive all of your material in the Edinburgh Research

Done

start

Inbox - Mic...

StOReMain

start

Inbox - Mic...

StOReMain

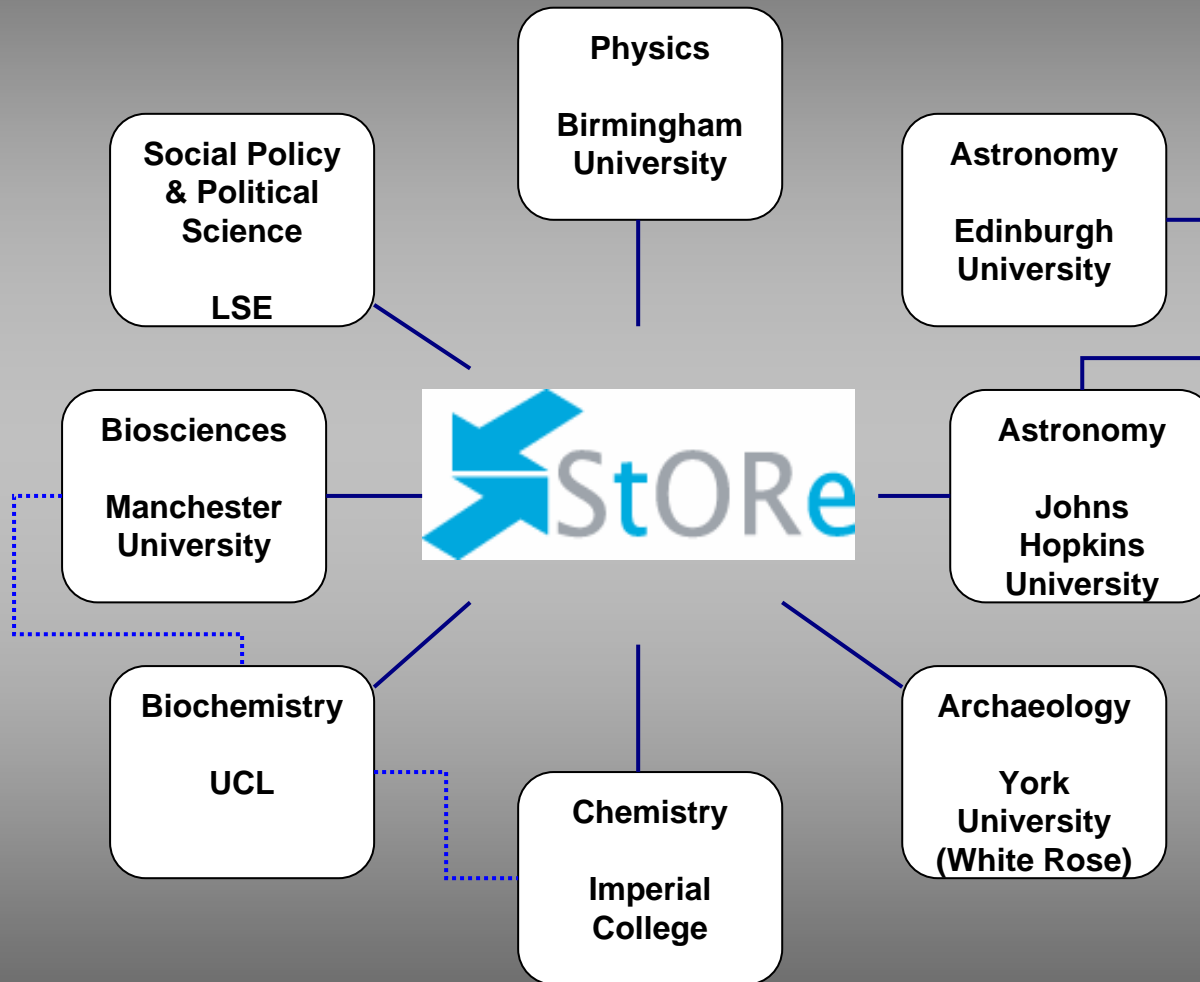
Microsoft P...

Edinburgh ...

Search Desktop

11:19

Libraries without walls 7



7 domains surveyed for data management and research process

Library origins of project – experience with generic tools and protocols

Survey team: library and domain expertise

Libraries without walls 7



New opportunities

- Explore a deeper level of detail
- Supplement published papers
- Validate experiments
- Track the use and improvement of research output
- Identify collaborators
- Confirm completeness of information searches

New risks

- Uncertainty of peer review
- Premature dissemination
- Subversion of scholarly paper
- Scavenging
- Lack of interpretative data



StORe survey

- 85% approved aims as likely to prove advantageous
- Questionable data management practices and a metadata conflict
- Tension between open access and data ownership



Data management

Common practice of storing unique and original research data on the hard drives of PCs and laptops (especially prevalent in chemistry and the biosciences)

“Once removed from the computers I hold the only available copies of most of my data”











Metadata – researchers identified

- The need for improved and universal standards
- A need for different metadata for different phases of the research lifecycle (raw, processed, published data and beyond)
- A clear link between the condition of metadata used and the level of support from information specialists

Libraries without walls 7



The assignment of metadata

I decide which terms to use and I assign them:		212
Research colleague(s) assign metadata on the team's behalf:		55
Research support staff assign metadata on the team's behalf:		22
Metadata are assigned by library/information services staff:		4
Metadata are assigned by the repository administrators:		37
Metadata are generated automatically:		63
It is not known who assigns metadata:		68
Other (please specify):		37

Libraries without walls 7



Types of support used	Academic staff	Research assistant	Post-graduate	Contract researcher	Independent researcher	Other	Totals
Documentary support	29	5	16	6	2	8	66
Support by intermediary	16	2	7	1	0	3	29
Repository-enabled support	56	5	7	5	1	9	83
No support is provided	60	7	23	4	3	9	106
Unknown	31	11	25	1	1	9	78
Other	5	1	4	1	0	3	14
No Answer	0	0	1	0	0	0	1
Totals	197	31	83	18	7	41	377

Libraries without walls 7



Assistance by library or IS staff	Academic staff	Research assistant	Post-graduate	Contract Researcher	Independent researcher	Other	Totals
Documentary (guidance notes etc.)	61	6	24	5	4	9	109
Formal training	38	3	27	3	0	3	74
Online or telephone Help	"...my university assigned a librarian to our department but I have not used her services..."					5	89
structuring searches						1	26
Help in the conduct of searches	20	4	11	3	0	4	42
Fully managed intermediation (searches and results)	3	2	5	0	0	1	11
Unknown	83	19	34	8	3	21	168



The research culture - a broad perspective

- Self-reliance and tradition of individual research endeavour
 - “No assistance required – it’s my responsibility”
 - “I rely on my own skills”
- Protection of research profile or status
 - “Data archives would become the preserve of the scavengers, unable to generate their own research ideas”
 - “The most important factors are those that would benefit my research profile”
- Conflict between a competitive environment and the aspiration to collaborate
 - “The competition for publication is intense and would not be aided by making data freely available”
 - “It should be a requirement for publicly funded research that data be made available”

Libraries without walls 7



Factors that **discourage** the sharing of research data

The threat of loss of ownership		202
Risks to an established research niche		104
Risk of premature broadcast of research findings		235
Subversion of intellectual property rights, including copyright		163
Ethical constraints relating to my research		58
Consideration of data protection and other confidentiality issues		115
The time/effort required to enable sharing		193
Risk of diversion from principal objectives through the generation of additional work		144
Risk to commercialisation opportunities		59
Increased competition for funding		77



The StORe team concluded that

- Data sharing is a fundamental principle of research, but...
- It is conducted mainly through private networking, with >50% through the personal exchange of e-mails or portable media
- Bureaucratic processes and organisations are anathema to researchers
- Both standards and simplicity are sought for research data management (esp. for metadata)
- Assisted self management is preferred to overt intermediation



Design for the StORe middleware

- Web 2.0 approach, similar to services like Flickr or MySpace, gives control to the researcher
 - Researchers determine which items are public / private
 - Researchers form collaborations with colleagues / 'friends'
 - Researchers select items for deposit and for publication
- Based on federations of institutional repositories and data archives
- Permanent links created between publications and underlying data
- Simple process for assignment of metadata
 - Searchable metadata assigned at collection level inherited by items within the collection
 - Collection owners add individual items plus minimum additional metadata (e.g. titles)

Libraries without walls 7



StORe
Source-to-Output Repositories

Home | Logout | About StORe | Administration

Search for in All Public Collections Search Adv. Search


All Collections | My Collections | Collaborative Collections | Approval queue | Profile

Item: Business Analysis

Collection: Edinburgh

View Item | Edit Item

Dublin Core

 [StOReBusinessAnalysis \(application/msword\)](#)

Add File or

EDIT

* Denotes required fields

*Title:

Creator: [Add Creator Field](#)

*Subject: [Add Subject Field](#)

*Subject:

A collaborative collection that I might use the pilot demonstration in Edinburgh

Home | A-Z index | Site map | Contact | Login | Search: Search text

ADD SUBJECT TAGS

The StORe pilot federation

Home | Help | Search | Site index | LSE for you

LSE Library

You are here - [Welcome to LSE](#) > [Library](#) > [LSE Research Online](#) > The purpose of institutional repositories in UK higher education : a repository manager's view

The purpose of institutional repositories in UK higher education : a repository manager's view

Rumsey, Sally (2006) The purpose of institutional repositories in UK higher education : a repository manager's view. *International journal of information management*, 26 (3). pp. 181-186. ISSN 0268-4012

Full text available as:



PDF - Requires [Adobe Acrobat Reader](#) or other PDF viewer.

Abstract

Open access publications may be a means of increasing availability of scholarly materials. However, there are other drivers which prompt HE institutions to spend staff time and investment on developing IRs. If the publishing arguments are put to one side, there is a separate and strong case for developing an IR which is justified by a library's functions of collecting, managing and curating materials.

Additional Information: Published 2006 © Elsevier B.V. LSE has developed LSE Research Online so that users may access research output of the School. Copyright and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users

Print-friendly page



About | Data | Support | R & D | News

SN 5278 -2001 Census: Special Licence Household Sample of Anonymised Records (SL-HSAR)

[Download/Order](#) 

External View

*** Internal User ***

[Documentation](#) [Variable List](#) [Keywords](#) [Publications](#) [Related Studies](#)

Title:
2001 Census: Special Licence Household Sample of Anonymised Records (SL-HSAR)
(2001 UK Individual Licensed Household SARs)

Subject Categories:
Census of Great Britain - Major studies
Population censuses - Population, vital statistics and censuses
Social indicators and quality of life - Society and culture

Depositor(s):

Accessing

Creating and

Finding
[Catalogue search](#)
[About catalogue](#)
[Help on searching](#)
[Browse by subject](#)
[Major studies](#)
[New releases](#)
[HASSET thesaurus](#)
[Other archives](#)

Libraries without walls 7



Source-to-Output Repositories

Search for

COLLECTION:

Madiera



E-MAIL

[Email Owner](#)

Creator	Ken
Subject	data portal, multilingual thesaurus, online data analysis
Description	Multilingual Access to Data Infrastructures of the European Research Area
Publisher	
Contributor	
Date	
Type	
Format	
Identifier	
Source	
Language	
Relation	
Coverage	
Rights	
Type of Research	Primary from Secondary (Creates new data based on existing data)
Data Repository	UKDA
Study	4995
Visibility	Private
Owner	Ken Miller (forum)
Contributors	

Libraries without walls 7



Source-to-Output Repositories

[Home](#) | [Logout](#) | [About StORe](#) | [Administration](#)

Search for in [All Public Collections](#) [Adv. Search](#)

[All Collections](#) [My Collections](#) [Collaborative Collections](#) [Approval queue](#) [Profile](#)

Item: Business Analysis

Collection: [Edinburgh](#)

[View Item](#) [Edit Item](#)

Dublin Core



StOReBusinessAnalysis
(application/msword)

[Add File or
Link](#)

[Move Item](#)

EDIT

* Denotes required fields

*Title:

Creator: [Add Creator Field](#)

*Subject: [Add Subject Field](#)

*Subject:

Description:

Publisher: [Add Publisher Field](#)

Contributor: [Add Contributor Field](#)

Date: [Add Date Field](#)

Type: [Add Type Field](#)

Format: [Add Format Field](#)

Identifier: [Add Identifier Field](#)

Source: [Add Source Field](#)

Language: [Add Language Field](#)

Relation: [Add Relation Field](#)

Coverage: [Add Coverage Field](#)

Rights: [Add Rights Field](#)

*Type of Research:

Study:

[Add Study Field](#)

ADD SUBJECT TAGS

books
crime
demonstrator
email
health
health subject
pilot
presentation
repositories data publications linkage
subject

Libraries without walls 7



Roles, Rights, Responsibilities, Relationships - 1

- Institutional Repositories tend to be managed by 'generalists' located in libraries and information services, and usually lacking data skills
- Subject repositories are staffed by domain experts with well-developed data handling skills
- Data sets tend to be unique
 - are deposited without multiple copies
 - are not self-describing and may
 - require interpretation, particular competencies and domain knowledge
- Representation information, which may include software programs, is required to assist with interpretation
- Closer partnerships between institutional and subject repository staff will be essential to building engagement with the designated community

Dealing With Data, Dr Liz Lyon, UKOLN, June 2007

Libraries without walls 7



Roles, Rights, Responsibilities, Relationships - 2

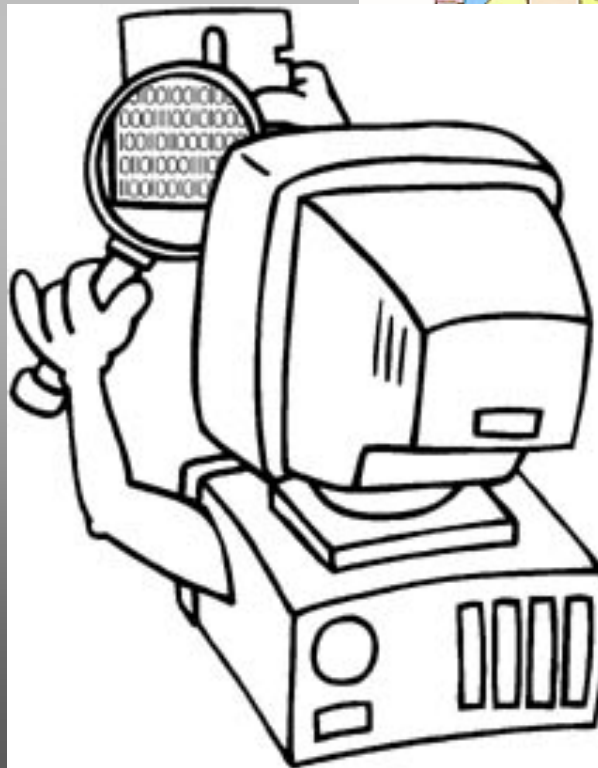
- Institutional repositories have a fundamental argument for sustainability because of their position within an institution
- They are frequently located in departments with a persistent service orientation - i.e. libraries or information services - whose main role is the management of information
- Many scholars care primarily about their discipline or domain, before their institution
- Scholars who have hybrid skills in domain science and data/information handling will be a valuable resource in the future

Dealing With Data, Dr Liz Lyon, UKOLN, June 2007

Libraries without walls 7

The next stage in the evolutionary growth of librarians...

Re-emergence of the scholar librarian?



Libraries without walls 7



Conclusion



Thank you